

情報分析技術

—大量の情報の中から価値ある情報を見つける—



風間 淳一 (かざま じゅんいち)

ユニバーサルコミュニケーション研究所
情報分析研究室 主任研究員

2004年に東京大学大学院を修了し、博士(情報理工学)を取得しました。その後、北陸先端科学技術大学院大学の助教を経て、2008年からNICTに勤務しています。専門は自然言語処理で、機械学習や確率モデルの自然言語処理への応用の研究を行ってきました。プライベートでは(最近はあまり行くことができませんが)釣りをしたり、けいはんな周辺の色々な名所を巡ったりしています。また、最近、家庭菜園を始めて、北陸先端大時代に知って好きになった加賀野菜の「金時草」を植えたりしています。

「大量の情報の中から価値ある情報を見つけるために私の所属する情報分析研究室で研究開発を行っている様々な技術、特に質問応答システム『一休』について紹介します。」

● はじめに

近年のインターネットの発達と、そこでの情報発信や検索の技術の発展には目覚ましいものがあります。しかし、現状で十分満足でしょうか。今でも、Web 検索の結果はページが単純にランキングされてリストアップされるだけで、検索する人が元々欲しかった情報を手に入れるためには、リストアップされたページを順に見ていくしか方法がありません。私の所属する情報分析研究室では、このような現状に満足せず、大量の情報の中から価値ある情報を見つけるための様々な技術の研究開発を行っています。

● 質問応答システム『一休』

質問応答とは、ユーザが「脳梗塞を防ぐのは何ですか?」のような質問文に対して、答えの候補を列挙する技術です。通常の Web 検索と大きく異なる点は、入力が文であることと、結果が「答え」の端的な列挙であるということです。このことから分かるように、通常の Web 検索よりも一段高度で、「価値ある情報の発見」という私たちの目的にかなった技術です。実は、質問応答の研究は古くから行われてきましたが、つい最近までは実用には届かない難しい技術ととらえられてきました。しかし、最近の技術の進歩で、かなり実用に近いところまで来ています。Web 上の大量のデータや、Wikipedia などのある程度組織化されたデータから、必要な知識を自動で抽出する技術が大きく進歩したためです。

我々の研究室でも、これまで開発してきた様々な技術を応用して、『一休』という質問応答システムを開発しています。このシステムの特徴は、Web に書かれていることであればおおよそどんなトピックに関する質問でも回答を返すことができるということです。音声入力機能を備え、スマートフォン上で動作します。技術的に

は、「言い換え認識」という技術が鍵となっています。私たち人間が使う言語では、同じ意味を表すために様々な表現を使うことができます。様々な言い方をされている所から回答を抽出するために、この言い換え認識の技術が重要となります。また、「推論」といういわば「計算機に考えさせる」技術を使って、Web に直接的に書かれていない回答を導くことも可能です。

ちなみに、我々がいるユニバーサルコミュニケーション研究所からほど近い京都府京田辺市にはとんちで有名な一休和尚(一休さん)が晩年を過ごしたお寺が有ります。『一休』という名前は、一休和尚のようにどんな難問でも答えられるようにという気持ちを込めてつけられました。

図 1 は、『一休』による質問応答の例を示した画面です。「脳梗塞を防ぐものは何ですか?」という質問に対して、「ダークチョコレート」「ナットウキナーゼ」「アスピリン」などの回答がされています。実は、「ダークチョコレート」という回答は、『一休』が用いている約 6 億ページの Web データの中にははっきりと簡単な形では書かれていません。『一休』は、Web 上の別々のところに書かれていた「ダークチョコレートは、動脈硬化を防止する働きがある。」「心臓病や脳梗塞などの原因になる動脈硬化は...」という 2 つの情報を自分で考えて組み合わせることで「ダークチョコレートは脳梗塞を防ぐ」という回答を導き出したのです。最近、アメリカの有名なクイズ番組で質問応答システムが人間のチャンピオンに勝ったというニュースがありました。それでも、そこで用いられたシステムには、『一休』のような進んだ推論の機能はありませんでした。『一休』は、この点で、一步進んだ質問応答システムと言うことができます。我々は、人間が行うような推論、つまり機械に考えさせる技術が、次世代の情報システムにおいて特に重要になると考えて、研究開発を進めています。

私も含めて、研究室の研究者は、開発のために日々『一休』のシステムを試していますが、通常の Web 検索では到底見つけることができない意外で価値ある情報を見つけ、それを自分の生活に実際に活かすという経験をしています。例えば、「味噌汁に何を入れますか?」という質問に対して「チーズ」を見つけて試してみたりといったことです。もう少し真面目な例を挙げると、「デフレの原因は何ですか」という質問に対して、『一休』は、大きな利益を上げているある大企業を意外な答えとして挙げました。Web でさらに調べると、その企業が莫大な利益を内部で貯金し社会に還元しないためであるという根拠を見つけることができ、またその後、雑誌に同主旨の記事が発表されました。これらの例で重要なことは、味噌汁の具としてチーズを、デフレの原因としてその企業を自ら

思いつくことはその分野の素人ではかなり難しいということです。もしかしたら、専門家でも難しいかもしれません。

このように見てくると、現在の通常のインターネット環境では、いかに多くのことがまだ可能になっていないかが分かります。私たちは『一休』の例で示したような、価値ある情報の発見を効率的に行うことができる技術を開発し、普及させたいと考えています。これにより、社会生活をさらに大きく進歩・進展させることができると考えています。

● 耐災害情報技術

2011年3月11日に発生した東日本大震災では、Twitterをはじめとする情報技術の可能性が強く認識された一方で、大量に飛び交う情

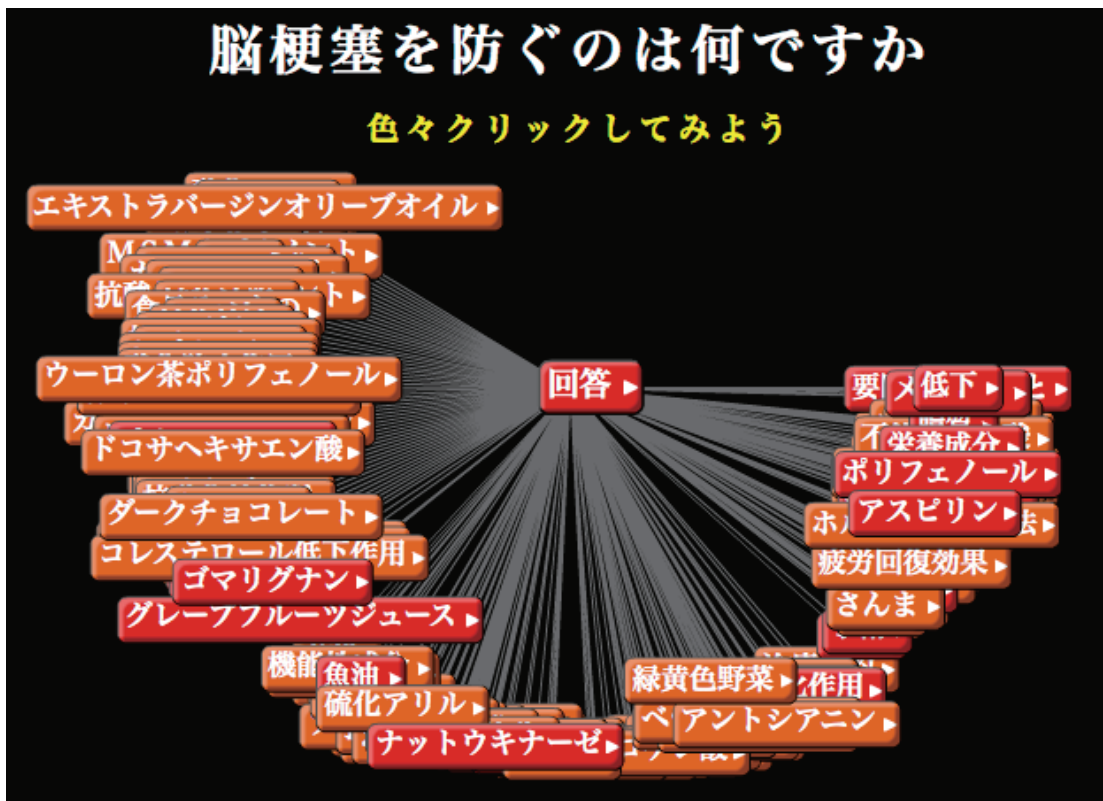


図1 「一休」による質問応答の例

報から状況を正確に把握し、救援や復興に役立てることがいかに難しいかを痛感させられました。結局のところ、私たちには「必要な情報を必要な人に届ける」技術の準備が十分にできていなかったのです。それが、大災害という極限状態になって露呈しました。あらためて、私たちが目標としてきた技術が必要とされていると感じています。情報分析研究室では、「宮城県で孤立しているのはどこですか?」といった災害時の質問に答えることが可能なシステムや、様々な事実や意見を分析して、多様な観点を考慮した冷静な判断を助けるシステムなどの開発を進めていきます(図 2)。

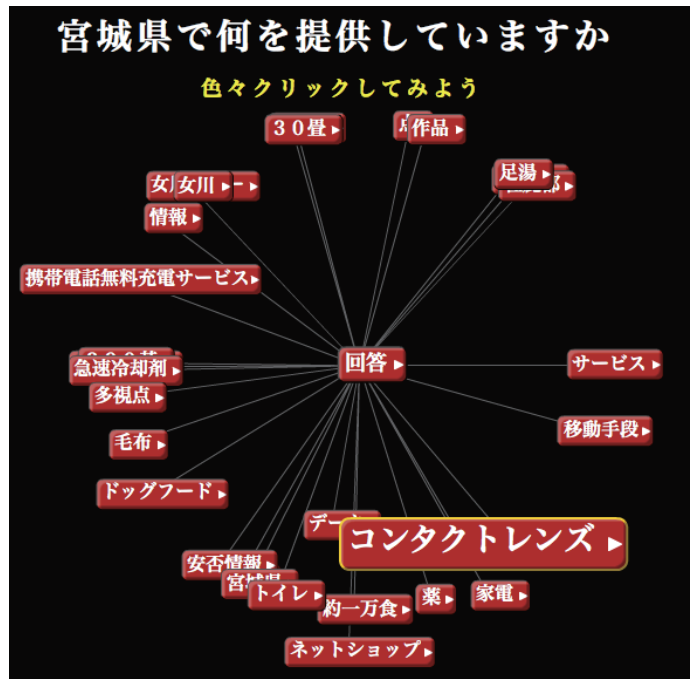


図2 「耐災害一休」プロトタイプによる質疑応答の例

● おわりに

今、日本は原発、デフレなど難しい問題が山積みですが、そうした問題の解決の道筋はいくらWeb 検索を使ったり、新聞を読んだりしてもなかなか分からないというのが多くの人の実感ではないでしょうか。現状の研究開発はまだ、そのレベルには達していませんが、将来的には、色々な人が書いている解決の道筋を可能な限り見つけ、分かりやすい形でまとめるといった技術にも繋げて行きたいと思っています。